THE PRIMER

JGI JOINT GENOME INSTITUTE DEPARTMENT OF ENERGY

spring 2016
volume 13
issue 1

Applying Genomics to Improve Cassava Breeding Strategies

Women peeling cassava for processing in Nigeria. (IITA)

Cassava is an easily cultivable staple crop for nearly a billion people around the world, and a primary source of calories. However, it is particularly vulnerable to plant pathogens, which can significantly reduce crop yields. To help improve breeding strategies for this root crop, a team led by researchers from the University of California, Berkeley and including researchers from the U.S. Department of Energy Joint Genome Institute (DOE JGI), a DOE Office of Science User Facility, have described cassava's genetic diversity ahead online April 18, 2016 in the journal *Nature Biotechnology*.

For the DOE, cassava's high starch content, ability to grow well on poor soil, and drought tolerance are among the reasons the root crop is of interest as a potential feedstock for biofuel production. As cassava roots contain 20–40% starch that costs 15–30% less to produce per hectare than starch from corn, in many parts of the world, particularly Africa and Southeast Asia, it represents a strategic source of renewable energy—biomass from which ethanol is being produced for transportation fuels. With the help of genomics, researchers hope to apply advanced breeding strategies that can improve *continued on page 5*

in this issue

Decoding Underlying Mechanisms: Notes from the 11th Annual DOE JGI User Meeting

"In nature, we have microbial communities collecting on thousands of different surfaces, and yet how communities are structured is where we have the greatest lack of understanding."

Speaking to a crowd at full capacity in the Walnut Creek Marriott, Dianne Newman of Caltech opened the 11th Annual DOE JGI Genomics of Energy & Environment Meeting on March 22, 2016. One of her messages, repeated by several speakers over the course of the meeting, is that much remains unknown about the underlying mechanisms and pathways of the systems being studied. Newman spoke about the importance of secondary metabolites, using a model system reliant on *Shewanella oneidensis* communities. "From the bottom up, we can begin to understand the rules of behavior, first from simple populations, to *continued on page 2*

Decoding Underlying Mechanisms

continued from page 1

"From the bottom up, we can begin to understand the rules of behavior, first from simple populations, to rules governing the more complex and vast microbial world."

–Dianne Newman

rules governing the more complex and vast microbial world."

Another theme of the meeting was the importance of having the right model system. In her closing keynote, Margaret McFall-Ngai from the University of Hawaii focused on the relationship between the bobtail squid and the luminescent bacteria *Vibrio fischeri*—one she's studied for 30 years—as an example of how experimental model systems reveal the principles underlying symbiosis. "It's my opinion model systems provide insights into mechanisms underlying symbiosis," she said, speaking of the power of simple systems for understanding more complex ones. Another powerful tool her team has exploited over the years is imaging: confocal microscopy, for example, helped them visualize the events associated with the onset of symbiosis in real time.

Manpreet Dhami of Stanford University uses nectar microbes, which she described as a relatively simple model system, for understanding interactions in community assembly. "If you arrive in a community patch first," she said, "you have a better chance of surviving because you can modify environment to your best use."

Rhona Stuart of Lawrence Livermore National Laboratory spoke about the model biofilm community she used in her postdoctoral studies to understand how and why cyanobacteria assimilate extracellular organic matter. With the help of nanoscale secondary ion mass spectrometry (NanoSIMS) imaging technology, she said, they were able to see that cyanobacteria have consumption preferences based on light and dark conditions, information that helped them understand the impact of nutrient availability.

Ben Cole, a DOE JGI postdoctoral fellow, spoke about how microbes colonize plant ecosystems. He focused on a single bacterium's interaction with the model plant *Arabidopsis thaliana*, talking about his group's interest in how plants initially acquire soil bacteria which may maintain plant health or improve crop yields.

Sarah Lebeis of the University of Tennessee discussed the project she's leading through the Facilities Integrating Collaborations for User Science (FICUS) program, a partnership between the DOE JGI and the Environmental Molecular Science Laboratory (EMSL) that allows researchers to combine the power of genomics and molecular characterization in one research project proposal. Her talk focused on characterizing the plant-growth promoting members of the duckweed microbiome. The duckweed genome was published by the DOE JGI for its potential as fast-growing biofuel feedstock that doesn't require the pretreatments that woody plants with high lignin content.

Kirsten Hofmockel from the Pacific Northwest National Laboratory spoke on another FICUS project, this one involving the challenge of understanding the microbial players—and their genes and genomes—that affect cellulose decomposition in soil.

Michelle O'Malley of UC Santa Barbara spoke about her team's efforts to study the applicability of gut fungi from zoo and livestock animals for breaking down plant biomass in commercial biofuel pipelines. Her findings are the first to be published from the DOE-EMSL FICUS partnership and are further detailed on page 7.

Kelly Wrighton of Ohio State University spoke about the need to understand the impact of hydraulic fracturing, colloquially known as fracking, on the microbial communities in the deep subsurface. Her team has found that the subsurface ecosystems are already being affected by the fracking process as microbes from the surface find their way underground, exploiting new habitats and nutrient sources. She is a collaborator through the DOE JGI Community Science Program as well as the FICUS program.

Orli Bahcall told the audience as an editor of the journal *Nature*, "what I really do is advocate for the best research." She talked about the changing nature of science communication, and how the journal is developing digital ways to tell stories and effectively package large collaborative projects.

Chris Bowler spoke about the massive datasets being generated as a result of the four-year Tara Oceans

expedition. His presentation was an update to the closing keynote delivered by Eric Karsenti in 2013. (Watch that presentation at http://bit.ly/JGI13_Karsenti.) Though 40,000 samples were collected, less than 600 have been processed and have already led to an ocean microbial reference catalog with 40 million genes, and the identification of more than 5,400 viruses, only 39 of which had been previously known. "The data deluge is coming," he warned the audience in a massive understatement.

Though his work isn't on scale yet with global oceans, Cornell University's Christopher Mason has global sequencing ambitions. His talk focused on studying city-scale microbiomes including subway systems to better understand the microbial diversity in shared ecosystems. The work could lead to previously unknown biosynthetic gene clusters "right under our fingertips," he pointed out. Building off his lab's work in swabbing and sequencing samples collected throughout the New York City subway lines, Mason said the team plans to go global next, developing metagenomic profiles for the 10 busiest subways around the world and a few dozen others.

Metagenomics, said Charles Chiu of UC San Francisco, could improve the rate at which disease-causing agents can be identified so that clinicians can produce real-time diagnoses and improve patient outcomes. He spoke about applying deep sequencing methods for medical diagnoses.

"Networks are typically hairballs, which are not informative," noted Christine Queitsch from the University of Washington as she talked about mapping regulatory DNA and transcription factor networks to track heat shock response and other stresses. In crops

Margaret McFall-Ngai

such as sorghum and rice, she said, understanding these networks could lead to improved breeding strategies.

Despite his midmorning slot, computational biologist Gene Myers filled the room as he gave an update on his work towards perfect *de novo* assembly. Best known for his roles in developing BLAST, as well as the shotgun sequencing protocol and the Celera assembler that helped speed the completion of the Human Genome Project, the former Berkeley Lab staff scientist (in the early 2000s) mused, "Wouldn't it be wonderful if we had every genome perfect so we'd never have to do it again?" His efforts to scrub the data "with as little loss of data as possible" for a perfect assembly can be tracked on his Dresden Azzembler blog, Dazzlerblog for short.

Several plant-related talks referenced DOE JGI collaborations or reference genomes available on the DOE JGI Plant Portal Phytozome (http://phytozome.jgi.doe.gov). For example, in order to develop new ways of growing plants, noted Jose Dinneny from the Carnegie Institution for Science, researchers need to understand how plants find water. He described an imaging system called Growth and Luminescence Observatory for Roots (GLO-Roots) that enables studies of root architecture, done with a team including the DOE JGI's John Vogel.

Jeanine Olsen from the Groningen Institute of Evolutionary Life Sciences described the international collaborative effort that led to the sequence of the first seagrass genome, that of the eelgrass *Zostera marina*. The work is described in more detail on page 6.

Jeffrey Ross-Ibarra of UC Davis discussed how plant genome size may vary based on adaptation, and may in turn impact population genetics. For example, maize on flat lands may have as much as 15% genetic variation, while teosinte growing on a hillside might have up to 30% genetic variation because of the altitude differences.

Elizabeth Kellogg from Donald Danforth Plant Science Center spoke about the comparative genomics of C4 plants, which include biofuel crops such as corn, sugarcane and switchgrass. As C4 photosynthesis helps these crops achieve high yields, "if we knew the C4 regulators, we could maybe engineer C3 plants to be more productive in a warming world," she said.

Tom Juenger from the University of Texas-Austin spoke *continued on page 4*

Decoding Underlying Mechanisms

continued from page 4

about his lab's work on finding genes underlying trait variation based on availability of nutrients such as water. The team studies *Panicum hallii*, a perennial grass related to the candidate bioenergy feedstock switchgrass, but with a shorter life cycle.

Tim Donohue from the University of Wisconsin-Madison and head of the Great Lakes Bioenergy Research Center gave an example of how the collaborations between the DOE JGI and the three Bioenergy Research Centers have led to genome-enabled discoveries. He spoke of the goal to redesign plant cell walls to make them more amenable for downstream processing. One example of this involves so-called "zip-lignin" poplars, in which the lignin was redesigned to have easily cleavable bonds.

Gene Myers and Ham Smith

"Wouldn't it be wonderful if we had every genome perfect so we'd never have to do it again?"

–Gene Myers

Mary Voytek, director of NASA's Astrobiology division, gave the audience a broader perspective, speaking about how genomics can help researchers understand the diversity of life on earth, while also providing insights into finding life on other planets. One example she gave involved lessons taken from analyzing 200 halophile genome sequences, and how this information might be applied to remotely sensed data suggesting the Mars subsurface has briny waters. "The work that you do is very important to the NASA astrobiology program," she said, mentioning plans to study "life as we do not know it" on Jupiter's moon Europa and Saturn's moon Titan.

Without leaving Earth, DOE JGI researchers have found microbes whose genetic codes diverge from a long-held prescribed vocabulary (see page 12). Within every genome sequence are three-letter codons that each represent one of the 20 regularly used amino acids, with three of the possible 64 three-letter codons reserved for stop signals. Robert Riley's talk focused on a clade of yeasts in which the 3-letter codon CUG is interpreted as the amino acid serine rather than the expected lysine, as well as another species that translates CUG to alanine.

Reshma Shetty from Gingko Bioworks talked about how the company has utilized several technology trends to source enzymes from publicly available systems for commercial gene synthesis. The abundance of genome sequence, improvements in mass spectrometry and the falling costs of gene synthesis allow the Gingko team to develop plant-inspired cultured aromas.

Nobel prize-winner Hamilton Smith of the J. Craig Venter Institute spoke about efforts to develop a living bacterial cell with a minimal, synthetic genome, work that was published in the journal *Science* just hours before his presentation. He revisited the original questions proposed by his team in 1995: "What is life, and what is the smallest number of genes needed to build a living cell?" The answer to the second question, as described in the *Science* paper, currently stands at 473 genes in a cell 531 Kb in size. Provocatively, 149 of those 473 genes (31.5%) cannot be assigned functions, even though they are required for growth.

A recap of the annual meeting as it happened is available at http://bit.ly/JGI2016storify. Selected videos from this meeting are available on the DOE JGI YouTube channel at http://bit.ly/JGI2016videos.

Cassava Diversity

continued from page 1

Members of the International Cassava Genetic Map Consortium (Left to right): Jessen Bredeson, UC Berkeley; Kahya Shuaibu, National Root Crops Research Institute (NRCRI), Umudike, Nigeria; Oluwafemi Alaba, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria; Cindy Ha, formerly of UC Berkeley and now at the University of Colorado; Jessica Lyons, UC Berkeley; Chiedozie Egesi, formerly of NRCRI and now with NEXTGEN Cassava at Cornell University; Simon Prochnik, DOE JGI; Ismail Rabbi, IITA, Nigeria. Bredeson, Ha, Lyons, Egesi, Prochnik and Rabbi are authors on the *Nature Biotechnology* paper.

cassava's resistance to diseases and improve crop yields.

The cassava genome was initially sequenced under the aegis of the DOE JGI Community Science Program and Roche 454 Life Sciences. Since the draft sequence was released in 2009, researchers have improved it with additional data in order to develop a chromosome-scale sequence, in part to apply the information toward improved breeding strategies. The cassava genome is available on the DOE JGI Plant Portal Phytozome (http://phytozome.jgi.doe.gov/).

In the paper, the team, which included UC Berkeley specialist Jessen Bredeson and postdoctoral scholar Jessica Lyons and DOE JGI's Simon Prochnik and Albert Wu, compared the cassava reference genome to the genomes of relatives castor bean (*Ricinis communis*), rubber tree (*Hevea brasiliensis*), Ceara rubber (*Manihot glaziovii*), and 53 cultivated and wild type cassava varieties from around the world. They found that the genetic diversity of cassava used in current breeding efforts has been greatly reduced in Africa, where viruses such as the cassava mosaic disease and the cassava brown streak disease have affected crop yields in many nations.

They were able to detect the genetic signature of past cassava improvement programs going back to the 1930's, which interbred cassava and Ceara rubber, and the persistence of these Ceara rubber regions in elite cassava varieties suggests they confer desirable traits. They also elucidated relatedness between many cultivated cassava varieties, which can help breeders maximize genetic diversity in improvement programs.

Steve Rounsley of Dow AgroSciences spoke about cassava genomics at the DOE JGI's 2014 Genomics of Energy and Environment Meeting. Watch the video at http://bit.ly/JGI14UMRounsley.

Collaborators on this project also included researchers at: International Institute of Tropical Agriculture (Nigeria and Kenya), National Root Crops Research Institute (Nigeria), Koronivia Research Station (Fiji); CIRAD (Vanuatu); Mikocheni Agricultural Research Institute (Tanzania); Naliendele Agricultural Research Institute (Tanzania); Donald Danforth Plant Science Center; Cornell University, Monash University (Australia); and Dow Agrosciences.

The work was funded by the Bill and Melinda Gates Foundation, the UK Department for International Development (DFID), and the NEXTGEN Cassava Breeding project.

Review the 2016 Community Science Program (CSP) Portfolio

A total of 27 projects were selected by the DOE JGI for the 2016 Community Science Program (CSP). The total allocation for the CSP 2016 portfolio is estimated to tap nearly 40 trillion bases (terabases or Tb) of the DOE JGI's plant, fungal and microbial genome sequencing capacity. The full list of projects may be found at http://jgi.doe.gov/our-projects/csp-plans/fy-2016-csp-plans/. Learn more about proposal calls for the Community Science Program at http://jgi.doe.gov/user-program-info/community-science-program/.

Seagrass Sequence Lends Salt Tolerance Insights

"Putting a genome together takes a big team effort," noted Groningen University's Jeanine Olsen in her talk at the DOE JGI's 2016 Genomics of Energy and Environment Meeting. Olsen led a consortium of 35 collaborators at 17 institutions around the world in a project to sequence the first seagrass genome, that of the eelgrass *Zostera marina*. The paper made the cover of the February 18, 2016 issue of *Nature*.

Zostera marina is the first marine flowering plant to be fully sequenced, work done through the DOE JGI's Community Science Program. As a foundational species in the coastal marine ecosystem, researchers are interested in understanding how the plant—and by extension other plants in the ecosystem—adapts to climate change.

Seagrasses are considered the "lungs of the sea," and coastal seagrass ecosystems account for an estimated 15% of carbon fixed in the global ocean, as well as impacting sulfur and nitrogen cycles. Jeremy Schmutz, head of the DOE JGI's Plant Program, emphasized that while eelgrasses are key players in coastal marine ecosystem functions, they are also endangered. "There are estimates that nearly a third of the eelgrass meadows worldwide have been destroyed by runoff into the ocean," he said, "reducing their potential capabilities as carbon sinks. Thus, studying the adaptive capacity of eelgrass is urgent to assist conservation efforts."

Despite the name, eelgrasses aren't true grasses but rather completely submerged marine flowering plants, or angiosperms, and members of an ancient monocot family. The team was interested in identifying the pathways that underwent major modifications upon *Zostera marina*'s return to the sea. To better understand the adaptations the plant made in returning to a saltwater environment, the team compared the eelgrass genome to its freshwater relative, Greater duckweed (*Spirodela polyrhiza*).

The team noted differences in genes related to cell wall structure due to adaptations to freshwater or terrestrial conditions. For example, plants such as duckweed have seemingly lost genes that help plants retain water in the cell wall, while eelgrass has regained these genes to better deal with osmotic stress at low tide.

"They have re-engineered themselves," said Olsen of the changes affecting the eelgrass cell walls. "Although this has been known biochemically for many years, the underlying pathway that produces these sulfated polysaccharides for the cell wall matrix, in combination with the expansion of low methylated pectins (zosterin), are now unravelled and their strong negatively charged nature is hypothesized to help protect the cells from osmotic stress. The *Zostera marina* genome is available on the DOE JGI Plant Portal Phytozome at http://phytozome.jgi.doe.gov/.

Jeanine Olsen's talk is at http://bit.ly/JGI2016Olsen. Learn more about this work at http://jgi.doe.gov/seagrass-genome-sequence-lends-insights-to-salt-tolerance/.

Harnessing Barnyard Bounty for Biofuels

Speaking at the 11th Annual DOE JGI genomics of Energy & Environment Meeting, UC Santa Barbara researcher Michelle O'Malley led off by crediting the collaborative science program offered by the DOE JGI and the Environmental Molecular Scientific Laboratory (EMSL), a national user facility at Pacific Northwest National Laboratory, for providing the resources that led to her team's recent *Science* paper.

"With the help of DOE, EMSL and JGI we've been able to crack open a difficult set of fungi with biotechnology relevance," she said. O'Malley, recipient of a DOE Office of Science Early Career Award within the Biological and Environmental Research Program and of a 2016 Presidential Early Career Award for Scientists and Engineers (PECASE), then described the work first published February 18, 2016.

O'Malley's team reported that anaerobic gut fungi found in the guts of goats, horses and sheep perform as well as the best fungi engineered by industry in their ability to convert plant material into sugars that are easily transformed into fuel and other products. The team's study is the first to result from a partnership between the two facilities called Facilities Integrating Collaborations for User Science or FICUS. The partnership allows scientists around the world to draw on capabilities at both user facilities to get a more complete understanding of fundamental scientific questions.

Companies want to turn biomass like wood, algae and grasses into fuel or chemicals. The problem: The matrix of complex molecules found in plant cell walls—lignin, cellulose and hemicellulose—combines to create the biological equivalent of reinforced concrete. When industry can't break down this biomass, they pretreat it with heat or chemicals. Or throw it away. Both options add to the cost of the finished product.

Anaerobic gut fungi colonize biomass, and secreted enzymes that release free sugars into their environment. (Artistic rendering of the fungi by UCSB engineering graphic designer Peter Allen)

Many farm animals have no trouble breaking down these same molecules, which inspired the research team to investigate. Their search started at the Santa Barbara Zoo and a stable in Massachusetts, where they collected manure from goats, horses and sheep. The fresher the sample, the better, for this barnyard bounty held live specimens of biomass-eating fungi.

"Nature has engineered these fungi to have what seems to be the world's largest repertoire of enzymes that break down biomass," said O'Malley.

O'Malley and her colleagues knew the fungi's hyphae excrete proteins, called enzymes, that break down plant material. Like tools in a toolbox, the more diverse the enzymes, the better the fungi can take apart plants and turn them into food. If industry can harness fungi with such a toolbox, it can more effectively break down raw biomass. The findings suggest that industry could modify the gut fungi so that they produce improved enzymes that will outperform the best available ones, potentially leading to cheaper biofuels and bio-based products.

Michelle O'Malley's talk can be viewed at http://bit.ly/JGI2016OMalley. For more information about the *Science* paper, go to http://jgi.doe.gov/biofuel-tech-fungi-straight-from-the-farm/.

Updating Microbial Diversity Depicted on the Tree of Life

Nearly 30 years ago, microbiologist Carl Woese radically proposed that the Tree of Life had a third branch—Archaea—rather than the two comprised of Eukarya and Bacteria. In a study published April 11, 2016 in *Nature Microbiology*, and with help from the DOE JGI, a team led by longtime DOE JGI collaborator Jill Banfield of UC Berkeley features a new depiction of the tree of life that better reflects the microbial diversity revealed through cultivation-independent techniques and bioinformatics methods. The team utilized the 30,437 eukaryote, bacterial and archaeal genomes publicly accessible through the DOE JGI's Integrated Microbial Genomes (IMG) database, and 1,011 newly-reconstructed genomes from previously uncharacterized lineages to reorganize the tree of life.

The expansion of the tree of life highlights the impact of advances in sequencing technologies that have allowed researchers to fill in previously unknown gaps of information. Additionally, the work emphasizes how much microbial diversity has been made accessible only through cultivation-independent techniques such as single-cell genomics and metagenomics. The updated tree of life builds off previous studies by the Banfield lab in which over 35 new groups or phyla of bacteria and nine new groups of archaea were identified. Many of the bacteria identified in the earlier study were sequenced as part of a 2015 DOE JGI Community Science Program (CSP) project that involved groundwater samples from a bioremediation site at Rifle, Colorado. This CSP project is also part of the Berkeley Lab Genomes-to-Watershed Scientific Focus Area (SFA), which involves over 50 scientists from Berkeley Lab and other institutions including UC Berkeley, Pacific Northwest National Laboratory, Colorado School of Mines, and Oak Ridge National Laboratory. With more accurate perspectives on microbial diversity and their relationships, scientists can make better inferences about adaptations to different environments, which can lead to discoveries of new enzymes and pathways to help address DOE missions in bioenergy and environmental processes.

An artistic representation of the tree of life, with the many groups of bacteria on the left, the uncultivable bacteria at upper right (purple), and the Archaea and eukaryotes (green)—which includes humans—at the lower right. (Graphic by Zosia Rostomian, Berkeley Lab)

Tracking Microbial Mat Formation in Yellowstone

Microbes such as those that thrive in the extreme environments of Yellowstone Hot Springs have been part of studies conducted at the DOE JGI for their potential bioenergy and environmental applications. A team of researchers from the Pacific Northwest National Laboratory and DOE JGI, led by longtime DOE JGI collaborator Bill Inskeep of Montana State University, developed a conceptual model that details how microbial mats are formed in hot, acidic springs in the Yellowstone caldera.

Microbial mats serve as model systems for studying microbial interactions and their influence over biogeochemical processes. Understanding how communities of microbes establish mats over time provides insights on how their environments determine the mechanisms they employ. The team sequenced DNA samples extracted from two acidic geothermal springs at

various timepoints over two months in Norris Geyser Basin at Yellowstone National Park. The data allowed the team to track the formation of microbial mats, beginning with primary colonization by *Hydrogenobaculum* species and *Metallosphaera yellowstonensis*, and how these populations as well as those of other microbes that colonized later changed over time in response to availability of nutrients such as oxygen and carbon. These studies continue to build on the decades of microbial field studies Inskeep and his team have done at Yellowstone National Park. The insights gained from this model, the team noted, could provide insights into microbial life at other hot springs ecosystems and, potentially, on other planets.

Multiple Methods for Microbial Diversity in One Lake

By sequencing the DNA collected at many times over nine years (2005–2013) from Wisconsin's Trout Bog Lake, a team led by University of Wisconsin scientists and including DOE JGI researchers was able to reassemble the genomes of dozens of freshwater microbes and follow how their genetic makeup changed over time. Surprisingly, they found different evolutionary processes at play within different microbial populations living in the same environment. For example, the team found one bacterial population undergoing a genome-wide selective sweep, a process predicted by the most prominent microbial evolutionary model (i.e., the ecotype model), but never before observed in the wild. During a genome-wide selective sweep, a single member of a population out-competes all others once it acquires an advantageous trait, thus purging genetic diversity within the population.

In contrast, they also found other bacterial populations that appear to experience gene-specific sweeps, indicating that high rates of genetic interchange within these populations preserved their diversity. During gene-specific sweep, only the small piece of DNA encoding the advantageous trait is shared with other members of the population through genetic recombination, thereby preventing a takeover by a single dominant strain. The team's results show for the first time that different populations coexisting in the same environment likely have dramatically different rates of genetic exchange, and that these differences allow populations to evolve according to different non-exclusive theoretical models. This is the first conclusion to come out of a project approved under the DOE JGI's Community Science Program in 2011 and was reported online January 8, 2016 in *The ISME Journal*. The paper has led to two opinion pieces from leaders in the fields of microbial evolution and population genomics.

A Window into Fungal Endophytism

Endophytes are organisms that reside within living plant cells without harming their hosts and often, in fact, contribute significantly to their survival and physiology. They can play roles not just in plant health, but also in carbon and nitrogen cycles. The evolution of endophytism and mechanisms by which endophytes interact with their hosts are still poorly understood. To gain insights into these questions, a team including DOE JGI researchers and longtime collaborators at Clark University sequenced the genome of *Xylona heveae*, an endophyte from a Peruvian rubber tree, and then compared its genome to the genomes of 36 related fungi within the phylum Ascomycota. The report was published in the January 2016 issue of *Fungal Biology*.

Peruvian rubber trees. (Marco Simola for the Center for International Forestry Research (CIFOR), CC BY-NC 2.0)

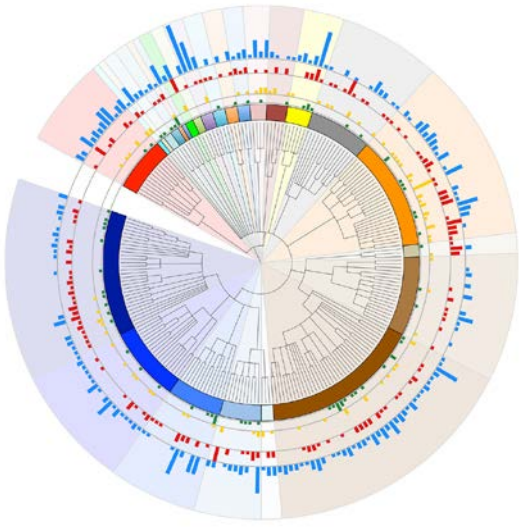
Based on this dataset of 37 fungal genomes, a third of which were sequenced and annotated at the DOE JGI, the team found that endophytes such as *X. heveae*, in order to avoid triggering their plant hosts' immune systems, adapted by reducing their CAZymes, cellulose-degrading genes that encode carbohydrate-active enzymes. At the same time, the fungal endophyte also appears to have expanded the number of enzymes that could boost its chances of survival within the host plant's intercellular spaces. The authors theorized that in *continued on page 10*

adapting to its host, *X. heveae* has reduced its wood-degrading capabilities to the point that it likely cannot switch to a different lifestyle nor become pathogenic. They also suggested that this fungal endophyte could be transmitted through insects. Understanding the evolution of the endophytic lifestyle could help researchers understand how plants and fungi developed symbiotic relationships, and how the mutualistic association provides host plants with beneficial traits.

Roles of DNA Methylation in Prokaryotes

The epigenome of a cell is the collected set of changes made to specific bases in its genomic DNA that affect how the genome is actually used and results from chemical modification (usually methylation). DNA methylation, the most common epigenetic change, is a process eukaryotes use to regulate gene expression; for example, keeping certain genes from turning on. Though prokaryotes (bacteria and archaea) are also known to have methylated DNA, the roles this process might play in these single celled organisms is less well understood. To learn more, a team including DOE JGI researchers relied on single-molecule, real-time (SMRT) sequencing at the DOE JGI and Pacific Biosciences to reveal DNA methylation patterns in 230 bacterial and archaeal genomes. They found evidence of DNA methylation in 215 microbes (93% of those sequenced). These data enabled the annotation of 600 enzymes that methylate DNA (MTases), a massive increase over known annotations.

While many DNA methylating enzymes are part of restriction modification systems, consistent with their known role in defense against phages and viruses, the findings suggest that a substantial number of others may be involved in genome regulation, and have a more crucial role in prokaryotic physiology and biology than had been previously suspected. By mapping and characterizing the epigenetic changes, scientists can associate those targeted genes with environmental adaptations and metabolic activities. Better understanding of such controls on gene expression and under what circumstances they are observed will improve the ability to predict when and where certain proteins are produced. In addition, this will inform how microbes interact with plants and other microbes involved in DOE mission interests such as plant bioenergy feedstock growth, advanced biofuel generation, and soil carbon processing. The report was published February 12, 2016 in *Plos Genetics* and was highlighted February 29, 2016 in *Nature Genetics Reviews*.



Phylogenetic tree of 230 sequenced organisms. (Figure from Blow et al. (2016) *Plos Gen.*)

A Signal in the Noise: Uncovering Kryptonia

In a study published January 27, 2016 in *Nature Communications*, DOE JGI researchers led a team that utilized a large collection of metagenomic datasets to uncover a completely novel bacterial phylum that they have dubbed "Kryptonia."

From 5.2 trillion bases (Terabases or Tb) of sequence in the Integrated Microbial Genomes with Microbiome Samples (IMG) system, the team identified long sequences that contained a phylogenetic marker (DNA corresponding to ribosomal RNA) commonly used to assign all life into a particular classification system that could not be placed into any recognizable phylum. Reconstructing the genomes from metagenomic datasets and single cell genomes yielded four lineages belonging to the novel candidate phylum, named Kryptonia (*Candidatus Kryptonia*) from the Greek word for "hidden."

"It's not every day that you find a completely new phylum," said DOE JGI's Emiley Eloe-Fadrosh. The DOE is seeking to uncover the true extent of the planet's microbial diversity in order to learn more about the genes, enzymes and metabolic pathways that play key roles in regulating critical biogeochemical cycles. More thorough surveys could lead to new strategies for DOE researchers to advance their energy and environmental investigations.

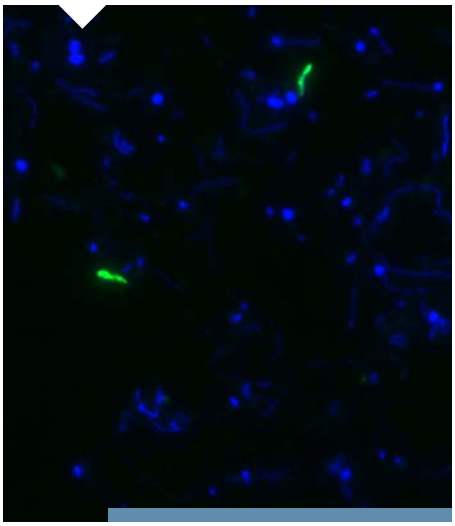
In a complementary article published February 1, 2016 in *Nature Microbiology*, DOE JGI researchers wondered how many other microbial lineages might have been missed, as Kryptonia had been, due to being overlooked by the Polymerase

Chain Reaction (PCR) primers most commonly used in these kinds of analyses. DOE JGI researchers employed datasets from the IMG system to determine how many microbial lineages might be missed from available sequences in existing datasets due to mismatches with the PCR primers currently used.

With help from the National Energy Research Scientific Computing Center (NERSC), the team analyzed more than 50,000 16S genes—gene sequences found in every microbe that can serve as an identifying marker—from metagenomic datasets and were able to find that as much as 10% of those sequences would be missed by the currently used PCR primers.

For details about both studies, go to http://jgi.doe.gov/uncovering-hidden-microbial-lineages-kryptonia-from-hot-springs/. Eloe-Fadrosh described the search for Kryptonia at the DOE JGI's 2015 Genomics of Energy and Environment Meeting at http://bit.ly/JGI15UMKryptonia.

A 'Ca. Kryptonia'-specific FISH probe was designed and used to visualize cells from Dewar Creek Spring sediment samples. 'Ca. Kryptonia' cells hybridizing with the probe are green, while other cells are visualized with 4',6-diamidino-2-phenylindole (DAPI; blue). (Composite image by Emiley Eloe-Fadrosh, DOE JGI)



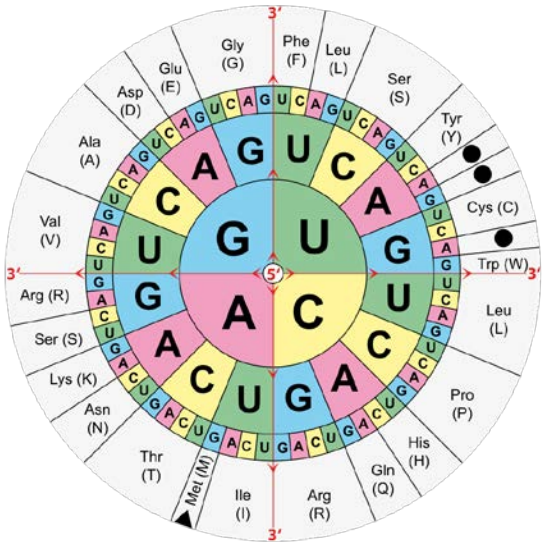
The Plasticity of the Genetic Code

Four letters—A, C, G and T—typically represent the four chemical bases that store information in DNA. The sequence of these same four letters genetically defines the genome, or "source code", for an organism. Within the genome sequence are three-letter blocks, called codons, that each represent one of the 20 regularly used amino acids, with three of the possible 64 three-letter codons reserved for stop signals. These amino acids are the building blocks of all the proteins that form the structures and carry out the myriad of functions necessary for cellular life. In some organisms, the three-letter codon UGA, which normally signals the end of a protein-coding gene, is hijacked to code for a rare genetically encoded amino acid called selenocysteine.

Published April 18, 2016 in the journal *Angewandte Chemie International Ed.*, DOE JGI and Yale University researchers discovered that microorganisms can recognize and use up to 12 more codons (that are usually used for "regular" amino acids) for selenocysteine. The finding adds credence to recent studies indicating that an organism's genetic vocabulary is not as constrained as was previously thought.

The work is a follow-up to two 2014 publications, including a *Science* paper by the DOE JGI group finding that some organisms interpret the three "stop" codons which terminate translation as amino acids. "Access to the tremendous resources at the JGI allowed us to quickly test challenging hypotheses generated from my research projects that have been supported over the long-term by DOE Basic Energy Sciences and the National Institutes of Health," said Yale professor Dieter Söll, lead author of the latest study. Delving into genomic data from uncultured microbes afforded researchers the opportunity to learn more about how microbes behave in their natural environments, which in turn provides information on their management of the various biogeochemical cycles that help maintain the Earth.

The findings, the team reported, "opens our minds to the possible existence of other coding schemes … Overall our approach provides new evidence of a limited but unequivocal plasticity of the genetic code whose secrets still lie hidden in the majority of unsequenced organisms." Details at http://jgi.doe.gov/proving-codon-genetic-code-flexibility/.



Starting from the four innermost letters and working to the outermost ring, this table shows which three-letter base sequence encodes which amino acid in the standard genetic code. (Wikimedia Commons)

JGI Director Eddy Rubin Steps Down

After 14 years guiding the DOE JGI from completing DOE's contributions to the Human Genome Project to transitioning the Institute into a National User Facility enabling the science of thousands of researchers focused on energy and environmental problems, Eddy Rubin stepped down as DOE JGI Director. Jay Keasling, Lawrence Berkeley National Laboratory (Berkeley Lab) Associate Laboratory Director for Biosciences, then announced that he asked DOE JGI's Axel Visel to serve as Interim Director effective March 24, 2016 up to the time that the permanent JGI Director is on board. The DOE JGI Director position description is posted at http://bit.ly/JGI-Director-PD.

Rubin will move to a position as Chief Scientific Officer for San Francisco-based startup Metabiota, a big data analytics company focused on infectious diseases and epidemic risk.

"I started out my career in medicine. Now that the Institute is running on all its scientific cylinders as a state of the art DOE Office of Science Genomic User Facility, I plan to return to these roots and focus my attention on addressing an important problem in human health," Rubin said. "I feel that the JGI is now in an outstanding position to expand its leadership in energy and environmental genomics. Equipped with a compelling 10-Year Strategic Vision and a great group of senior staff, the DOE JGI is ready to take on the exciting new opportunities associated with our planned move to the Berkeley Lab main campus and into the new Integrative Genomics Building."

After serving his residency at the University of California, San Francisco (USCF) and a genetic fellowship and faculty position there, Rubin moved from UCSF to Berkeley Lab in 1989, becoming a Senior Scientist in 1992 and Group Leader for Human Genome Biology in 1997. He took over as the DOE JGI Interim Director in 2002 and was appointed Director of the Institute and Director of the Berkeley Lab Genomics Division in January 2003.

"Equipped with a compelling 10-Year Strategic Vision and a great group of senior staff, the DOE JGI is ready to take on the exciting new opportunities associated with our planned move to the Berkeley Lab main campus and into the new Integrative Genomics Building."

–Eddy Rubin

SAVE THE DATE

12th Annual Genomics of Energy & Environment Meeting

March 20–23, 2017
Walnut Creek, California

http://bit.ly/JGI-UM12

Who should attend?

All current Community Science Program (CSP) investigators and collaborators, as well as those considering an application for future CSP calls. We also welcome any and all researchers and students interested in energy and environmental genomics.

Topics:

Microbial genomics, fungal genomics, metagenomics, and plant genomics; genome editing, secondary metabolites, pathway engineering, synthetic biology, high-throughput functional genomics, high-performance computing applications and societal impact of technological advances. State-of-the-art presentations by invited speakers as well as short talks selected from poster abstracts.

Contact The Primer
David Gilbert, Managing Editor
DEGilbert@lbl.gov
Massie Santos Ballon, Editor

PAC 16-JGI-2830